An explanatory model for food-web structure and evolution

A. G. Rossberg*, H. Matsuda, T. Amemiya, K. Itoh

Yokohama National University, Graduate School of Environment
and Information Sciences, Yokohama 240-8501, Japan

*Corresponding author. Tel.: +81-45-339-4369, fax: +81-45-339-4353
E-mail addresses: rossberg@ynu.ac.jp (A.G.R.), matsuda2@ynu.ac.jp (H.M.),
amemiyat@ynu.ac.jp (T.A.), itohkimi@ynu.ac.jp (K.I.)

Abstract

Food webs are networks describing who is eating whom in an ecological community. By now it is clear that many aspects of food-web structure are reproducible across diverse habitats, yet little is known about the driving force behind this structure. Evolutionary and population dynamical mechanisms have been considered. We propose a model for the evolutionary dynamics of food-web topology and show that it accurately reproduces observed food-web characteristic in the steady state. It is based on the observation that most consumers are larger than their resource species and the hypothesis that speciation and extinction rates decrease with increasing body mass. Results give strong support to the evolutionary hypothesis.

Keywords: food-webs, evolution, networks, model validation

1   Introduction

The complex networks of feeding relations in ecological communities, food webs, have fascinated researchers for over thirty years (Cohen et al., 1990). Using carefully collected empirical data, previous indications (Cohen et al., 1990; Havens, 1992) could be confirmed that several aspects of the structure of empirical food webs are reproducible across habitats as diverse as Caribbean islands, deserts, and lakes (Camacho et al., 2002b; Milo et al., 2002; Garlaschelli et al., 2003; Williams and Martinez, 2000; Cattin et al., 2004). Most striking, perhaps, were findings that certain descriptive model food webs, constructed by following just a few simple rules for connecting the elements of an abstract species pool, can quantitatively reproduce large sets of basic statistics of food-web topology (Cohen et al., 1990; Williams and Martinez, 2000; Cattin et al., 2004). These models do not yet expose the dynamical processes or mechanisms by which the structures they describe are formed. But they set a standard by which the accuracy of any dynamical explanation should be measured (Williams and Martinez, 2000).

Two major mechanisms for shaping food webs have been considered. One is the selection of those food webs that support stable population dynamics, because the degree of stability observed in nature is hard to achieve with simple, randomly assembled food-web models (May, 1973; Yodzis, 1981). The other mechanism is evolutionary dynamics, which is suggested by the observation that the similarities in the trophic role of species and their phylogenetic kinship are highly correlated (Cattin et al., 2004). Even though there has been plenty of work on dynamical food-web models incorporating population dynamics, evolutionary dynamics, or combinations of both (for a review, see Drossel and McKane, 2003), non of these models could so far be shown to reproduce food-web structure to the accuracy of the more recent descriptive models, namely the *niche model* (Williams and Martinez, 2000) and the *nested hierarchy model* (Cattin et al., 2004). Purely evolutionary models (e.g., Amaral and Meyer, 1999) have been developed mainly for understanding the dynamics of the total number of species, and not the detailed food-web structure. Work on models combining evolutionary and population dynamics [e.g. Caldarelli et al. (1998); Drossel et al. (2001); Yoshida (2003)] has not yet arrived at conclusive results, one problem being high model complexity. A recent systematic tests of a plain population dynamical model by Montoya and Solé (2003) turned out clearly negative.

Here we introduce a new model (called *speciation model* below) for the evolutionary dynamics of food webs, and show that its accuracy is comparable to the descriptive models. For an in-depth mathematical analysis of the model we refer to Rossberg et al. (2005).

2   Model definition

We model the evolutionary dynamics of an abstract pool of species belonging to a given habitat, and the topology of the network of feeding relations between them. In order to separate the plain macroevolutionary dynamics from the complex conditions that determine fitness, evolution is modeled as being undirected. This simplification is here not meant to be a statement about nature, but is used as a technique to isolate the effect of particular processes. Population dynamics, for example, does not enter the model.

The speciation model is most naturally described in terms of a continuous-time stochastic process. (See Appendix A for a description of our computational implementation.) Its dynamics consists of two parts, the dynamics of the community and the dynamics of the food web connecting the species. Table 2 lists the model parameters.

2.1 Evolution of the species pool

Only three kinds of macro-evolutionary processes are taken into account: speciations and extinctions within a habitat and adaptations of new species to the habitat. Anagenesis (evolution without speciation) is omitted for simplicity. Since changes in the environment can be fast on evolutionary time scales, it is for this model more appropriate to conceive a habitat as a particular set of environmental conditions supporting an ecological community, rather than just a specific location.

Species at higher trophic levels tend to be larger than at lower levels (Memmott et al., 2000; Leaper and Huxham, 2002). Due to this correlation, a possible correlation between body size and speciation- and extinction rates leads to correlations between evolution rates and trophic level. Arguments in favor of both, evolution rates increasing (Fenchel, 1993) and decreasing (Bush, 1993) with body size have been put forward. Empirically, the question seems to be undecided. Generally one would assume that evolution rates are on the average proportional to some power M^α of the adult body mass M. We note that $\alpha = 0$ is just as difficult to motivate as any other choice. Based on the comparison of the observed distributions of the number of links to resources and consumers (Camacho et al., 2002b), it can be argued that $\alpha < 0$ is more plausible in the context of the model proposed here (Rossberg et al., 2005).

We associate each species i with a "speed parameter" s_i (triangles in Fig. 1) that determines the magnitude of its evolutionary rates $\sim \exp(s_i)$. The community dynamics is characterized by the values of the rate constants for new adaptations r_1, speciations r_+, and extinctions r_-, the range $[0, R]$ of allowed values for s, and the evolutionary speed/size dispersion D. The complexity of the processes driving evolution is accounted for by modeling extinctions, speciations, and adaptations of new species to the habitat as independent random events: For any short time interval $[t, t + dt]$, the probability that a species i becomes extinct is $r_- \exp(s_i)\, dt$ and the probability that it speciates is $r_+ \exp(s_i)\, dt$. When i speciates, a new species j with slightly different speed parameter $s_j = s_i + \delta$ is introduced into the community, where δ is a zero-mean Gaussian random variable with $\operatorname{var} \delta = D$. If $s_i + \delta$ exceeds the range $[0, R]$, $s_j = -(s_i + \delta)$ or $s_j = 2R - (s_i + \delta)$ are used instead. The probability that a new species with a speed parameter in a small range $[s, s + ds]$ adapts to the habitat and joins the community is $r_1 \exp(s)\, ds\, dt$ for all s between 0 and R.

2.2 Dynamics of the trophic network

The food web is described by the topology of the network of directed trophic links connecting species. Parameters determining its dynamics are the loopiness λ, the raw connectivity C_0, and the reconnection probability β. Following Cohen et al. (1990), Williams and Martinez (2000), and Cattin et al. (2004), we take the fact into account that consumers are typically larger than their resource species. Empirical data show that they are, parasites excluded, at least not much smaller in terms of adult body mass (Memmott et al., 2000; Leaper and Huxham, 2002). We assume that there is an upper limit h for the ratio of the adult body sizes of resource and consumer species. Resource species are at most h times larger than their consumers. Making use of the assumption that evolution rates $\sim \exp(s)$ scale as M^α with $\alpha < 0$, the difference of the speed parameter s of a resource and its consumer has an upper limit approximately given by $\Lambda := |\alpha| \times \ln h$. The degree by which the trophic ordering is violated within the species pool is expressed by the loopiness parameter $\lambda = \Lambda / R$. Summarizing, size constraints restrict the set of *possible consumers* of a species i to the species l with $s_l < s_i + \lambda R$ and the *possible resources* to the species h with $s_h > s_i - \lambda R$ where $0 \le \lambda \le 1$ (shaded area in Fig. 1). Thus, cannibalism and trophic loops (Polis, 1991) can occur.

When a new species k is adapting to the habitat, a possible consumer of k becomes its

3

consumer with probability C_0 and, with the same probability, a possible resource of k becomes its resource. When a species i speciates (Fig. 1), the connections of the descendant species j are determined in three steps. First, all possible resources and consumers of i that are also possible resources and consumers of j are linked to j. Possible resources and consumers of j that are not possible consumers or resources of i are linked with probability C_0. Then, for each possible consumer or resource of j, the information whether it is connected is discarded with probability β. Finally, a link is established with probability C_0 to any consumer or resource for which the information was discarded. Species that are both possible resources and possible consumers of j are treated like two species: a consumer and a resource. This procedure ensures that the average ratio of possible links to realized links remains at C_0, in line with the premise that evolution is essentially undirected.

The steady-state properties of the model are independent of the initial conditions. Approximations for some basic properties can be obtained analytically (Rossberg et al., 2005). For example, the expected density of species has a constant value $r_1/(r_- - r_+)$ along the s axis. Thus, the observed broad distributions of logarithmic body masses (e.g. Fenchel, 1993) are here modeled by a simple boxcar function. The number of species S is on the average $\overline{S} = Rr_1/(r_- - r_+)$ and, with L denoting the number or links, the average directed food-web connectivity defined as $C = L/S^2$ is approximately $\overline{C} \approx C_0 \left(1 + 2\lambda - \lambda^2\right)/2$. The fraction of species that entered the habitat by speciation is r_+/r_-, the rest $(1 - r_+/r_-)$ entered as new adaptations.

3 Model verification

The speciation model was compared with seven of the best available empirical webs (see Table 5), all of which exclude parasites, in line with our model. Collecting food-web data is not at all an easy task, and the data is sometimes criticised for being incomplete or inconsistent. But it was for this kind of data that the regularities mentioned in Section 1 have been found, which suggests that the data does carries substantial information characteristic for food webs.

A simple, visual comparison of the connection matrix of the food web of Little Rock Lake (Martinez, 1991) with a model sample (Fig. 2) shows that the model reproduces the characteristic mixture of randomly scattered blocks and isolated links found in empirical data. The empirical web has several omnivorous species of high trophic level, which leads to a high density of links near the left edge of the connection matrix. Such isolated, dense blocks are only occasionally found in simulations. They are not typical for empirical webs either.

A systematic model verification was done based on twelve commonly used food-web statistics (see Appendix B). We kept the parameters $R = \ln 10^4$ and $D = 0.0025$ fixed, since their precise value has little effect (doubling either D or R changes the χ^2 values computed below by less than one.), and set $r_- = 1$ without loss of generality. Depending on r_+, λ, and β (see below), parameters r_1 and C_0 were chosen so that estimates of $\overline{S}$ and $\overline{C}$ after data standardization (Appendix B) match with the species abundance S_e and connectivity C_e of the empirical webs. Maximum likelihood estimates (see Appendix C) assuming a multivariate Gaussian distribution for the twelve food-web statistics were used for the remaining model parameters r_+, λ, and β (Table 5). As a benchmark, food-web statistics were also computed for the niche model (Williams and Martinez, 2000) and for the nested hierarchy model (Cattin et al., 2004). Neither has free parameters. As is shown in Fig. 3, there is good overall agreement between empirical and model data for all models.

As a systematic measure for the goodness of fit, we computed the χ^2 statistics given by

$$\chi^2 = (\mathbf{v}_e - \overline{\mathbf{v}})^T \mathbf{C}^{-1} (\mathbf{v}_e - \overline{\mathbf{v}}), \tag{1}$$

where $\mathbf{v}_e$ denotes the vector of the twelve food-web statistics for an empirical web, and $\overline{\mathbf{v}}$ and $\mathbf{C}$ are the corresponding model mean and covariance matrix (Appendix C). In terms of χ^2 statistics (Table 5) the speciation model (χ^2_S) is more accurate than the niche (χ^2_N) and the nested hierarchy model (χ^2_H), even when differences in the number of statistical degrees of freedom (DOF = number components of $\mathbf{v}$ minus number of fitting parameters) are taken into account. For example, the overall p value for either niche or nested hierarchy model is less than 10^{-24} times that of the speciation model. In all but one case, the speciation model yields the lowest χ^2. This is partially due to inherently larger model variances and partially due to a better accuracy of the model averages: in 46 out of 84 cases the average of the speciation model is closest to the empirical value ($p < 0.001$). Smaller webs are modeled particularly well. For the five food webs with S_e up to 35 (after species lumping) all χ^2_S values lie within the 5% confidence interval $\chi^2_S < 16.9$; and so does their sum $\sum \chi^2_S = 50.5 < 61.7$.

4   Discussion

In order to exclude the fallacies that the surprisingly good quantitative fit of the speciation model (1) does not actually depend on evolutionary processes or (2) is merely a consequence of the data standardization and evaluation procedures deployed, the model was also evaluated with $r_+ = 0$ fixed, i.e. with all speciation processes inhibited. In this case the parameter β also does not have any effect, so that the only remaining fitting parameter is λ. The model fits worsened considerably. We obtained χ^2 values ranging from 65 (Chesapeake Bay) to over 1000 (Ythian Estuary, Little Rock Lake).

Of course this does not prove that there is no other explanation for the observations, but it shows that, for the good fit of the model, the speciation events are crucial. Two effects of the evolutionary dynamics are probably particularly important: (1) It leads to approximately exponential distributions in the number of resources of a species (Rossberg et al., 2005), distributions that are similar to those obtained with the niche model by prescribing a beta distribution for the "niche width" Camacho et al. (2002a,b). (2) Related species have similar sets of consumers and resources, an effect that the algorithm of the nested hierarchy model (Cattin et al., 2004) was set up to mimic.

The speciation model is at least as accurate in its predictions as the niche model, which was itself a big improvement over the, by now historical, cascade model (Cohen et al., 1990). The speciation model is the first that involves a mechanistic explanation, confirming the earlier hypothesis (Cattin et al., 2004) that the structure of food webs can be understood as the effect of a sequence of evolutionary events. Recent theoretical and empirical results (McCann, 2000) show that the population dynamics in complex food webs can be stabilized by mechanisms such as weak links and adaptive foraging (Kondoh, 2003). This indicates that population-dynamical restrictions on food-web structure are less severe than one might first assume, making room for evolutionary dynamics to leave its traces.

5   Acknowledgements

Appendices

A   An Efficient Implementation of the Speciation Model

An evaluation of the speciation model based on a straightforward discretization of speed parameters s and time t would require excessive amounts of computation time. Here the data structures and algorithms that were used in our, more efficient, implementation are described.

A.1   From Poisson processes to event sequences

The main loop of our algorithm advances the dynamics of the food webs by one evolutionary event at each iteration. The question which event comes next is decided based on the following considerations: Call the quantity $\sigma_i := \exp(s_i)$ the *raw evolution rate*, or simply *evolution rate* of species i. We first note that the probability for new species with an evolution rate in the range $[\sigma, \sigma + d\sigma]$ to adapt is

$$r_1 \exp(s) \, ds \, dt = r_1 \, \sigma \, \frac{ds}{d\sigma} \, d\sigma \, dt = r_1 d\sigma \, dt. \tag{A.1}$$

Thus, the total rate at which species adapt to the habitat is simply

$$\nu_1 := \int_{\sigma_{\min}}^{\sigma_{\max}} r_1 d\sigma = r_1 \left(\sigma_{\max} - \sigma_{\min} \right) \tag{A.2}$$

with $\sigma_{\min} := \exp(0) = 1$ and $\sigma_{\max} := \exp(R)$. Since all possible events are assumed to be statistically independent, the probability that *any* new-adaptation, extinction, or speciation event occurs in the infinitesimal time interval $[t, t + dt]$ is $\nu_{\text{tot}} dt$ with

$$\nu_{\text{tot}} := \nu_1 + \left(r_- + r_+ \right) \sum_i \sigma_i, \tag{A.3}$$

where the sum runs over the indices of all species currently in the species pool. The time τ to the next event follows the distribution

$$P(\tau) = \nu_{\text{tot}} \, \exp(-\nu_{\text{tot}} \tau). \tag{A.4}$$

If the probability for some event X to occur in an interval $[t, t + dt]$ is $\nu_X \, dt$, then the probability for event X to occur as the next event is

$$\frac{\nu_X}{\nu_{\text{tot}}}. \tag{A.5}$$

To see this, consider first the situation with only two kinds of events: event A occurring at rate ν_A and event B at rate ν_B (both are Poisson processes). The joint probability distribution for the time τ_A to the next event A and the time τ_B to the next event B is

$$P_{AB}(\tau_A, \tau_B) = \nu_A \, \exp(-\nu_A \tau_A) \times \nu_B \, \exp(-\nu_B \tau_B). \tag{A.6}$$

The probability that A occurs before B is

$$\begin{aligned}
P\left[\tau_A < \tau_B\right] &= \int_0^\infty d\tau_A \int_{\tau_A}^\infty d\tau_B P_{AB}(\tau_A, \tau_B) \\
&= \int_0^\infty d\tau_A \nu_A \exp(-\nu_A \tau_A - \nu_B \tau_A) \\
&= \frac{\nu_A}{\nu_A + \nu_B}.
\end{aligned} \tag{A.7}$$

The probability (A.5) is obtained by letting event A be event X and event B be any other event.

For the sake of formal simplicity the range of allowed values for the speed parameter s is reduced from $[0, R]$ to the half open interval $[0, R[$ without affecting numerical results. In order to decide which event comes next in the sequence of events, a random number ρ equally distributed in the range $[0, \nu_{tot}[$ is generated. If $\rho < \nu_1$, the next event is taken to be a new adaptation and in view of Eq. (A.1) the evolution rate of the new species is chosen as a random number equally distributed in the range $[\sigma_{min}, \sigma_{max}[$. If $\nu_1 \leq \rho$, the next event is taken to be either the extinction or a speciation of the species i that satisfies

$$\left(r_- + r_+\right) \sum_{\sigma_j > \sigma_i} \sigma_j \leq \rho - \nu_1 < \sigma_i + \left(r_- + r_+\right) \sum_{\sigma_j > \sigma_i} \sigma_j. \tag{A.8}$$

Another random number is used to decide if the species i goes extinct [probability $r_-/(r_- + r_+)$] or speciates.

A.2   Data structures for keeping track of the network

The connectivity between species is stored in a matrix of bits $\{b_{ij}\}$ with $b_{ij} = 1$ is j eats i and $b_{ij} = 0$ otherwise. In order to avoid having to re-arrange the full matrix b_{ij} with every event, the number of rows and columns of $\{b_{ij}\}$ is chose much larger than the actual number of species is the pool. Indices are dynamically allocated with every new adaptation or speciation, and deallocated upon extinction. Unused indices are kept in a stack data-structure. The determination of the connectivity follows directly the model definition.

A.3   Sampling the steady state

It is important to notice that the main loop of the algorithm, which simulates one evolutionary event per iteration, does not proceed with constant speed along the time axis t. Episodes with many species require more iterations per unit time than episodes with few species. In order to sample the steady state evenly in t, the an estimate $\hat{t}$ of t has to be used, since the true time between events is left undetermined with the method described above. Such an estimate can be obtained by starting with $\hat{t} = 0$ and incrementing $\hat{t}$ by the current expectation value of the time between events $\tau_{tot} := \nu_{tot}^{-1}$ at each iteration.

The portion of the food web that involves small species with large σ reaches the steady state much faster than the portion that involves only large, slow species (small σ). In the initial stage of the simulation, it is therefore sufficient to account for the portion involving only slow species. In practice, this is achieved by setting the effective value R' of the parameter R to a low value R_{start} (e.g., $R_{start} = 1$) at the beginning of the simulation and increasing R' after each iteration by a small amount proportional to τ_{tot}. When R' has reached R, the simulation is continued for some time, and then a steady-state food web is sampled. For the next steady-state sample, the complete food web is discarded, and the simulation is restarted with a $R' = R_{start}$. We started each run with an empty species pool.

In order to check if the relaxation time is long enough to reach the steady state, we verified that, for the steady-state samples taken, the density of species along the s axis has the theoretically expected constant value $r_1/(r_- - r_+)$ for $D = 0$.

B   Food-web properties

The food-webs properties that were used to characterize and compare empirical and model webs were: the clustering coefficient (Camacho et al., 2002b; Dorogovtsev and Mendes, 2002) (*Clust* in

Figure 3); the fractions of cannibalistic species (Williams and Martinez, 2000) (*Cannib*) and species without consumers (Cohen et al., 1990) (*T*, top predators); the relative standard deviation in the number of resource species (Schoener, 1989) (*GenSD*, generality s.d.) and consumers (Schoener, 1989) (*VulSD*, vulnerability s.d.); the web average of the maximum of a species' Jaccard similarity (Jaccard, 1908) with any other species (Williams and Martinez, 2000) (*MxSim*); the fraction of triples of species with two or more resources, which have sets of resources that cannot be ordered to be all contiguous on a line (Cattin et al., 2004) (*Ddiet*); the average (Cohen et al., 1990) (*aChnLg*), standard deviation (Martinez, 1991) (*aChnSD*), and average per-species standard deviation (Goldwasser and Roughgarden, 1993) (*aOmniv*, omnivory) of the length of food chains, as well as the $\log_{10}$ of their total number (Martinez, 1991) (*aChnNo*), with the prefix *a* indicating that these quantities were computed using the fast, "deterministic" Berger-Shor approximation (Berger and Shor, 1990) of the maximum acyclic subgraph (MAS) of the food web. The number of non-cannibal trophic links not included in the MAS was measured as *aLoop*. When the output MAS of the Berger-Shor algorithm was not uniquely defined, the average over all possible outputs was used.

The statistics were computed after standardization of food webs to allow a consistent comparison of data: Since in many records of empirical food webs the lowest trophic level is poorly resolved, all species without resource species were lumped to a single "trophic species" after dropping disconnected species and before the usual lumping of trophically equivalent species (Cohen et al., 1990), for both empirical and model data.

C Maximum Likelihood Estimates, Goodness of Fit, and Confidence Intervals

Maximum likelihood estimation of model parameters and a systematic characterization of the goodness of fit ideally require knowledge of the joint probability distribution of the food-web statistics in the steady state for any given set of model parameters. We used the methods described below to obtain satisfactory approximation in a computationally feasible way.

Since species number S and directed connectivity C tend to fluctuate strongly in the steady state of the speciation model, it is impracticable to restrict sampling to those webs with S and C in close vicinity of the corresponding empirical values S_e and C_e as done by Williams and Martinez (2000); Cattin et al. (2004). Instead, all webs with values of S and C differing by less than 30% from S_e and C_e were used for the statistical analysis. In order to estimate the steady-state mean $\overline{\mathbf{v}}$ and covariance matrix $\mathbf{C}$ of the twelve food-web statistics *conditional* to $S = S_e$, $C = C_e$, we first computed steady-state mean $\overline{\mathbf{u}}$ and covariance matrix $\mathbf{D}$ of the vector

$$\mathbf{u} = \begin{pmatrix} \mathbf{v} \\ \mathbf{w} \end{pmatrix}, \tag{C.1}$$

consisting of the twelve food web statistics $\mathbf{v}$ and the vector $\mathbf{w} = (S, C)^T$ from $N = 1000$ samples within the $\pm 30\%$ range. Using $\mathbf{w}_e = (S_e, C_e)^T$ and components of $\overline{\mathbf{u}}$ and $\mathbf{D}^{-1}$ written in the block matrix forms

$$\overline{\mathbf{u}} = \begin{pmatrix} \overline{\mathbf{u}}_{\mathbf{v}} \\ \overline{\mathbf{u}}_{\mathbf{w}} \end{pmatrix}, \quad \mathbf{D}^{-1} = \begin{pmatrix} \mathbf{M}_{\mathbf{vv}} & \mathbf{M}_{\mathbf{vw}} \\ \mathbf{M}_{\mathbf{wv}} & \mathbf{M}_{\mathbf{ww}} \end{pmatrix}, \tag{C.2}$$

estimates for $\overline{\mathbf{v}}$ and $\mathbf{C}$ can be obtained as

$$\overline{\mathbf{v}} = \overline{\mathbf{u}}_{\mathbf{v}} + \mathbf{M}_{\mathbf{vv}}^{-1} \mathbf{M}_{\mathbf{vw}} \left(\overline{\mathbf{u}}_{\mathbf{w}} - \mathbf{w}_e \right), \tag{C.3a}$$

$$\mathbf{C} = \mathbf{M}_{\mathbf{vv}}^{-1}. \tag{C.3b}$$

The linear projection (C.3) is straightforwardly verified to be exact when $\mathbf{u}$ and, as a result, $\mathbf{v}$ have multivariate Gaussian distributions and $N \to \infty$. Maximizing the corresponding likelihood function for the empirical statistics conditional to $S = S_{\mathrm{e}}$, $C = C_{\mathrm{e}}$,

$$\frac{1}{\sqrt{(2\pi)^{12}\,|\mathbf{C}|}}\exp\left[-\frac{1}{2}(\mathbf{v}_{\mathrm{e}} - \overline{\mathbf{v}})^T \mathbf{C}^{-1}(\mathbf{v}_{\mathrm{e}} - \overline{\mathbf{v}})\right], \tag{C.4}$$

is equivalent to minimizing

$$\chi^2 + \ln|\mathbf{C}| \tag{C.5}$$

with χ^2 as given by Eq. (1) in the main text. The maximum likelihood estimates for the model parameters r_+, λ, and β were obtained by minimizing (C.5) using the global optimization algorithm of Schonlau et al. (1998).

When the number of samples N used for estimating $\overline{\mathbf{v}}$ and $\mathbf{C}$ in the formula for χ^2 (main text) is small, statistical errors of these estimates broaden the probability distribution of χ^2 beyond the usual χ^2-distribution. In the simplest case χ^2 (precisely $N\chi^2$) then follows Hotelling's T^2 distribution, for which analytic expressions are known. We decided to use the infinite sample-size confidence intervals (in 95% of all cases $\chi^2 < 16.9$ for 9 DOF and < 61.7 for 9×5 DOF) instead of the finite sample size intervals for $N = 1000$ (< 17.1 and < 65.1). This is the more conservative choice and, since our procedure for estimating $\overline{\mathbf{v}}$ and $\mathbf{C}$ is different from the one assumed for Hotelling's T^2 distribution, it is not clear if these finite-sample-size corrections would still apply.

References

Amaral, L. A. N., Meyer, M., 1999. Environmental changes, coextinction, and patterns in the fossil record. Phys. Rev. Lett. 82 (3), 652–655.

Baird, D., Ulanowicz, R. E., 1989. The seasonal dynamics of the Chesapeake Bay ecosystem. Ecol. Monogr. 59, 329–364.

Berger, B., Shor, P. W., 1990. Approximation algorithms for the maximum acyclic subgraph problem. In: Proceedings of the First Annual ACM-SIAM Symposium on Discrete Algorithms. SIAM, Philadelphia, pp. 236–243.

Bush, G. L., 1993. A reaffirmation of Santa Rosalia, or why are the so many kinds of small animals? Academic Press, Ch. 14, pp. 229–249.

Caldarelli, G., Higgs, P. G., McKane, A. J., 1998. Modelling coevolution in multispecies communities. J. Theor. Biol. 193, 345.

Camacho, J., Guimerà, R., Amaral, L. A. N., 2002a. Analytical solution of a model for complex food webs. Phys. Rev. E 65, 030901R.

Camacho, J., Guimerà, R., Amaral, L. A. N., 2002b. Robust patterns in food web structure. Phys. Rev. Lett. 88 (22), 228102.

Cattin, M.-F., Bersier, L.-F., Banasek-Richter, C., Baltensperger, R., Gabriel, J.-P., 2004. Phylogenetic constraints and adaptation explain food-web structure. Nature 427, 835–839.

Cohen, J. E., Briand, F., Newman, C. M., 1990. Community Food Webs: Data and Theory. Vol. 20 of Biomathematics. Springer, Berlin, and references therein.

Dorogovtsev, S. N., Mendes, J. F. F., 2002. Evolution of networks. Advances in Physics 51 (4), 1079–1187.

Drossel, B., Higgs, P. G., McKane, A. J., 2001. The influence of predator-prey population dynamics on the long-term evolution of food web structure. J. Theor. Biol. 208, 91–107.

Drossel, B., McKane, A. J., 2003. Modelling food webs. In: Bornholdt, S., Schuster, H. G. (Eds.), Handbook of Graphs and Networks: From the Genome to the Internet. Wiley-VCH.

Fenchel, T., 1993. There are more small then large species? OIKOS 68 (2), 375–378.

Garlaschelli, D., Caldarelli, G., Pietronero, L., 2003. Universal scaling relations in food webs. Nature 423, 165–168.

Goldwasser, L., Roughgarden, J., 1993. Construction of a large Caribbean food web. Ecology 74, 1216–1233.

Hall, S. J., Raffaelli, D., 1991. Food-web patterns: lessons from a species-rich web. J. Anim. Ecol. 60, 823–842.

Havens, K., 1992. Scale and structure in natural food webs. Science 257, 1107–1109.

Jaccard, P., 1908. Nouvelles recherches sur la distribution florale. Bull. Soc. Vaudoise Sci. Nat. 44, 223–270.

Kondoh, M., 2003. Foraging adaptation and the relationship between food-web complexity and stability. Science 299, 1388–1391.

Leaper, R., Huxham, M., 2002. Size constraints in a real food web: predator, parasite and prey body-size relationships. OIKOS 99, 443–456.

Martinez, N. D., 1991. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. Ecol. Monog. 61, 367–392.

May, R. M., 1973. Stability and complexity in model ecosystems. Princeton Univ. Press.

McCann, K. S., 2000. The diversity-stability debate. Nature 405, 228–233.

Memmott, J., Martinez, N. D., Cohen, J. E., 2000. Predators, parasitoids and pathogens: species richness, trophic generality and body size in a natural food web. Journal of Animal Ecology 69, 1–15.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U., 2002. Network motifs: Simple building blocks of complex networks. Science 298, 824–827.

Montoya, J. M., Solé, R. V., 2003. Topological properties of food webs: from real data to community assembly models. Oikos 102 (3), 614–622.

Polis, G. A., 1991. Complex desert food webs: an empirical critique of food web theory. Am. Nat. 138, 123–155.

Rossberg, A. G., Matsuda, H., Amemiya, T., Itoh, K., 2005. Some properties of the speciation model for food-web structure, `http://arXiv.org/q-bio.PE/0502007`, under review.

Schoener, T. W., 1989. Food webs from the small to the large. Ecology 70, 1559–1589.

Schonlau, M., Welch, W., Jones, D., 1998. Global versus local search in constrained optimization of computer models. In: Flournoy, N., Rosenberger, W. F., Wong, W. K. (Eds.), New Developments and Applications in Experimental Design. Vol. 34. Hayward, pp. 11–25.

Warren, P. H., 1989. Spatial and temporal variation in the structure of a freshwater food web. Oikos 55, 299–311.

Williams, R. J., Martinez, N. D., 2000. Simple rules yield complex food webs. Nature 404, 180–183.

Yodzis, P., 1981. The stability of real ecosystems. Nature 289, 674–676.

Yoshida, K., 2003. Evolutionary dynamics of species diversity in an interaction web system. Ecological Modelling 163, 131–143.

Table 1: Empirical food webs, fitting parameters, and goodness of fit

Food Web	key	$S_\mathrm{e}^{\,\mathrm{a}}$	$C_\mathrm{e}^{\,\mathrm{a}}$	$\dfrac{r_+}{r_-}$	r_1	λ	C_0	β	$\chi_\mathrm{N}^{2\,\mathrm{b}}$	$\chi_\mathrm{S}^{2\,\mathrm{c}}$	$\chi_\mathrm{H}^{2\,\mathrm{d}}$
								number of statistical DOF:	11^{e}	9	12
Bridge Brook Lake (Havens, 1992)	BB	15	0.28	0.91	0.17	0.12	0.37	0.059	43	13.9	44
Skipwith Pond (Warren, 1989)	Sk	25	0.32	0.93	0.21	0.009	0.53	0.012	83	11.4	64
Coachella Desert (Polis, 1991)	Co	27	0.34	0.96	0.13	0.006	0.58	0.014	37	3.9	152
Chesapeake Bay (Baird and Ulanowicz, 1989)	Ch	27	0.08	0.96	0.21	0.25	0.06	0.029	14	7.5	23
St. Martin Island (Goldwasser and Roughgarden, 1993)	SM	35	0.14	0.80	0.92	0.000	0.23	0.034	17	13.8	12
Ythan Estuary (Hall and Raffaelli, 1991)	Yth	78	0.06	0.95	0.67	0.001	0.08	0.040	58	36.8	82
Little Rock Lake (Martinez, 1991)	LR	80	0.15	0.99	0.13	0.025	0.16	0.006	52	24.2	165

[a] after species lumping (Appendix B)
[b] niche model (Williams and Martinez, 2000)
[c] speciation model (this work)
[d] nested hierarchy model (Cattin et al., 2004)
[e] the statistic $Ddiet$ (Appendix B), which is always zero for the niche model, was excluded

Table 2: Model parameters and their theoretical range

model parameter		range
rate constant for new adaptations	r_1	$r_1 > 0$
rate constant for speciations	r_+	$r_+ \geq 0$
rate constant for extinctions	r_-	$r_- > r_+$
total range of evolution rates	R	$R \geq 0$
rate/size dispersion constant	D	$D \geq 0$
loopiness	λ	$0 \leq \lambda \leq 1$
raw connectivity	C_0	$0 \leq C_0 \leq 1$
re-connection probability	β	$0 \leq \beta \leq 1$

[newSpeciation2.eps]
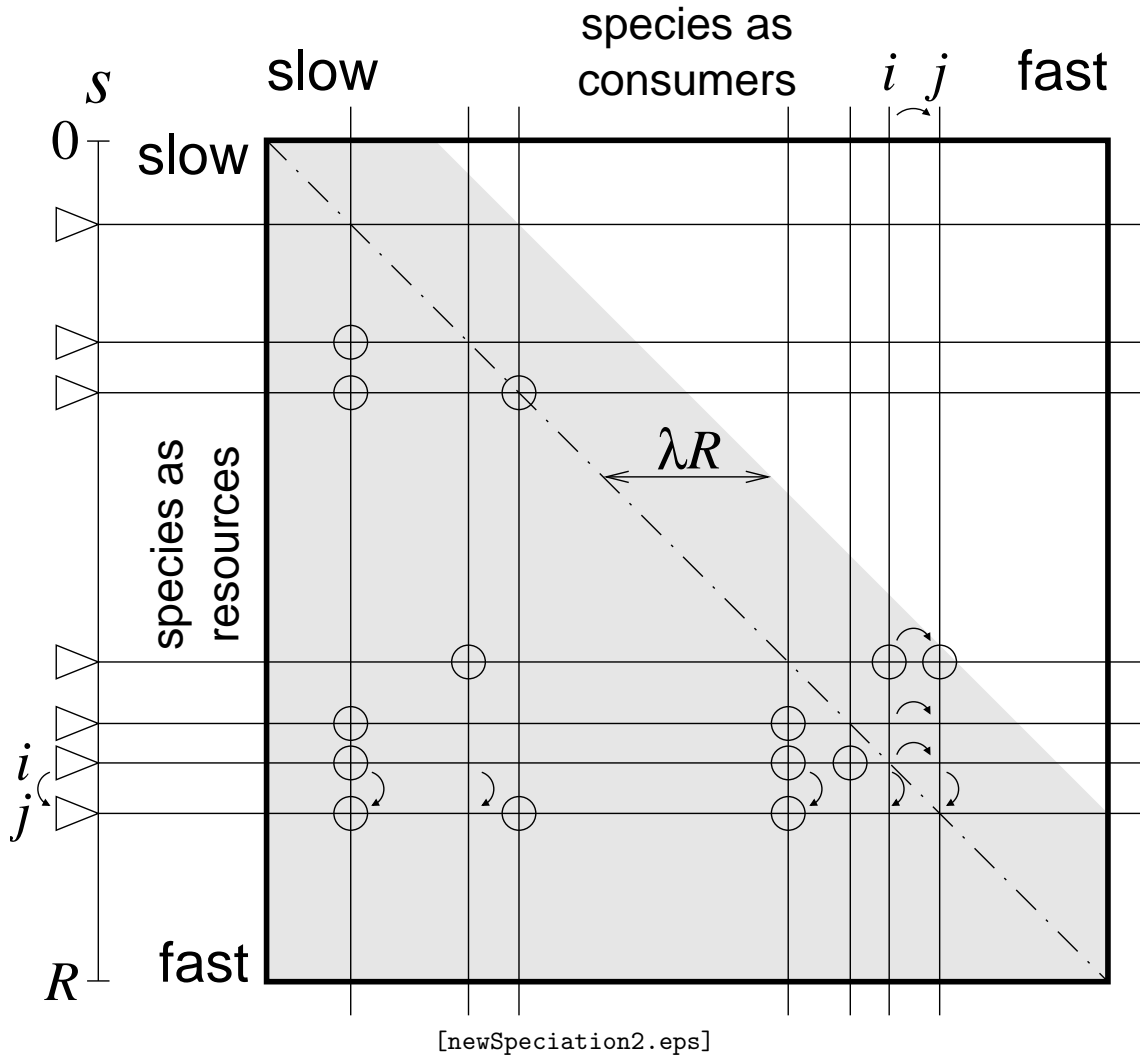
Figure 1: Illustration of the speciation model

The food web is represented by a connection matrix (large square). Horizontal lines represent species as resources, vertical lines the same set of species as consumers. Trophic links are indicated by circles at the intersection points. The s-axis marks the evolution rate of each species. The process shown is a speciation. Species j evolves from species i. Most trophic links of j are copied from i, but some are modified.

Little Rock Lake   Simulation Sample

```
X...............X...                              ...................................................
....................                              ...................................................
X...............X...                              XX..................................................
X...............X...                              XX..................................................
X...............X...                              XX..................................................
X.XX................                              XX..................................................
..XX................                              XX..................................................
X.XXX..X.......X....                               XX..................................................
X.XXX..X.......X....           .X......X.          XX..................................................
X.XXX..X.......X....           .X......X.          XX..................................................
X.XXX..X....               .X......X.              XX..................................................
X.XXX..X....                                       XX..................................................
X..X....X..X..               X......                XX..................................................
X.XX...X.X.                                        XX..................................................
..X....X.X.                                        XX..................................................
X.XX....X.X.                                       XXX.X..X...X...X...X.
..XX...X.X.                                        XX..X...X..X.
X.XXX..XXX.                                        XX..X...X..X.
X..XX..XXXX.                                       ......X..X.
..X.X..XXXX.                                       XXX.X..X...X...X...X.
X.XXX..XXXX.                                       ......X.........X.
...XX..XXXX.                                       XXX.X..X.......X.
..XXX..XXXX.                                       X.X.X..X...X.
X.XXX..XXXX.........X.                              X.X.X..X...X...X.
..XXXX.XXXX.                                        X.X.X..X...X...X.
X..X..XXXX.                                         ...........X.
X.XXX..X.......X..........X......X.                  XXX.X..X.......X...X...X.
..XXX..XXXX......X.                                 ....X.X...X...XX.X...X.
....X..XXXXX.............X.                          ..X.................X.
..X..X..X.................X.                         .X...XX...X.......X.X..X.......XX...X......XXX.
X.XX...XXX................X.                         ..X.....................X.
X.XXX..XXXX..............X.                          ..X.................X.
..X.X..XXXXXX........X...X...........X.              ..X...XX...X.......X.X..X.......X.X...X....X...XXX.
X..X...XXXX.............X.                           X.X.X..X..X..X.....X...X.
.XXXX.XXXX...........X.                              X.X.X..X..X..X.....X.X.
.XXX..XXXX........X..X...X..X..X.                    X.X.X..X..X..X.....X.X.
.XXXXXXXXX...X..X...X...X.X.                         X.X.X..X..X..X.....X.X.
.XXXXXXXXX...X..X...X...X.X..X.X.X.                  X.X.X..X..X..X.....X.X.
...XX..XXXXXX....X..X...X.X.                         X.X.X..X..X..X.....X.X.
.XXXXXXXXX...X..X...X...X.X.X.X.X.                   X.X.X..X..X..X.....X.X.
...X...XXXX...........X...X...X..X.X.X.....X.        ..X...XX...X.......X.X..X.......XX...X......XXX.
.XXXXXXXXX...X..X...X...X.X.X.X.                     ..X...XX...X.......X.X..X.......XX....X....X..X.
...X..XXXXX...........X...X..X.X.X.X.....X...X.      ..X...XX...XX......XXXX.X......XXX.....X....X..X.X.
.XXXXXXXXX...X..X...X...X.X.X.                       XXX....X........X................X.
..XX.XXXXXX..X..X...X...X.X..X.X.X.                  XXX....X.X..X........................X.
..XX.XXXX....X..X...X...X.X..X.X.X.                  XXX....X........X................X.
..XX.XXXXXX..X..X...X...X.X..X.X.X.                  XXXX...X.X...X.X.......X.X.X..........X.
..XX...XXX.............X.........X.X.                ..X...X..X...X.......X.X..X....XXX...X....X..XXX.
..XX...XXXX........X.........X.X.                    ..X.X....X..XX....XX...X......X............X.X.
.XX.XXXXXX...X..X...X...X..X.X.X.X.                  ..X......X....X.X.XX...X...X.......X......X.X.
..X.X...XXXX...........X.X.X...X.X.X.X.              ..X......XX..XX....X...X...........X......X.X.
X.XXX.XXXXX..................X.X.                    ..X.X......XX..XX...X...X....X.X.X......X.X.
.....................X..........X........X.          ..X.X......XX..XX...X...X...X.X.X.X.....X.X.
X.XXXX.XXXX...X..X..........X..X.X.X.                ..X...XX...X.......X.X..X.......XXX...X....X..XXX.
....X.X..XX..X.......X.....X........X.X.             ..X.X....X.........X..................X...X.
..X.X...XXXX..........X.X...X..X.X.X.X......X..X.    X.X.X.X....X........................X....XX.
....XXX..X..X..X...X...X.X..XX.X.X.X......X.         X.X.X.X....X...................X....XXX.
.....XXXXXX..X.X.X..X...X..X.X..X.X.X.X.......X.X.   ..X...XX...X.......X.X..X.......XXX...X....X..XXX.
.XXX..XXXXXX......X.X..X...X.X..X.X.X.X.....X.       ..X...XX...X.......X.X..X.......XXX...X....X..XXX.
...XX.X..XX...............X..X.X..XX.X.X.X.X.....X.X. X.X.X.X....X.........X....XXX.....X....X.
...XX.X..XX.X............XX...X.X..XX.X.X.X.X...X.   ..X...X........X........X....XXX...X....X..XXX.
...X.X..XX..............X..XX...XX.X.X.X.X....X.     XXX....X.........X....XXX...X....X..XXX.
..X..XXXXXX..........X.X....X..X.X.X......X.         XXX....X........X....XXX...X....X..XXX.
...XX.X..XX...........X...X..X.X.X.X.X.X.....X.X.    XXX....X.X......X....XXX...X....X..XXX.
..X.XX.X..XX.X......XX...X.X.X.X.X.X.X.X.....X.      XXX....X........X....XXX...X....X..XXX.X.
...X.X..XX..............X..X.X.X.X.X.X.X.X.....X.    XXX....X........X....XXX...X....X..XXX..X.
.XXXX.X..XX.X.........X.X..X.X.X.X.X.X.X.....X.      ..XX....X..X...X....X.X.X.XX...XXXXX.XX..X........XX...XX.X.
...X..XXXXX...............X.X..X.X.X.X.X....X.       ..XX....X..X...X....X.X.X.XX...XXXXX.XX..X........XX...XX.X.
...XX..XXXX...........X...X..X.X.X.X.X.......X.      ..XX....X..X...X....X.X.X.XX...XXXXX.XX..X........XX...XX.X.
...X..XXXXX..........X...X..X.X.X.X.X.....X.         ..X....X...XX.X.XX....X...X........X....X............XXX.
...X.XXXXX.............X...X..X.X.X.X.X....X.        ..X....X...XX.X.XX....X...X........X....X...........X.X.
..X.X..XXXXX....X...X......XX..X.X.X.X.X....X.       ..X....X...XX.X.XX....X...X........X....X...........X.X.
....XX.X..XX.X..X....XX...X.X..X.XXXXX.X.X...X..X.X. X.X.X.X.....X........................X....XX..............X.
...X.X.XX.X.....X......XX...X.X..X.X.X.X.X.....X.X.  XXXXX.XXX.XXXXXXX.XXXXX.XXXXXXXXX...XX.XXXXXXXXXXXXXXXXXXXXX.X.X.X..XXXXXX.
....X.X.XXX.X.....X......XX...X.X..X.X.X.X.X...X.X.
....X.X..XX.X...X......XX...X.X..X.XXXXX.X.X....X.X..X.X..X.X.
.XXXX.XXXXXX....X..X..X.X..X.X..XX..X.XXX.......XX..XXXX.XX.XXXXX.XX.XX.X.XXXXXXXXX.
```
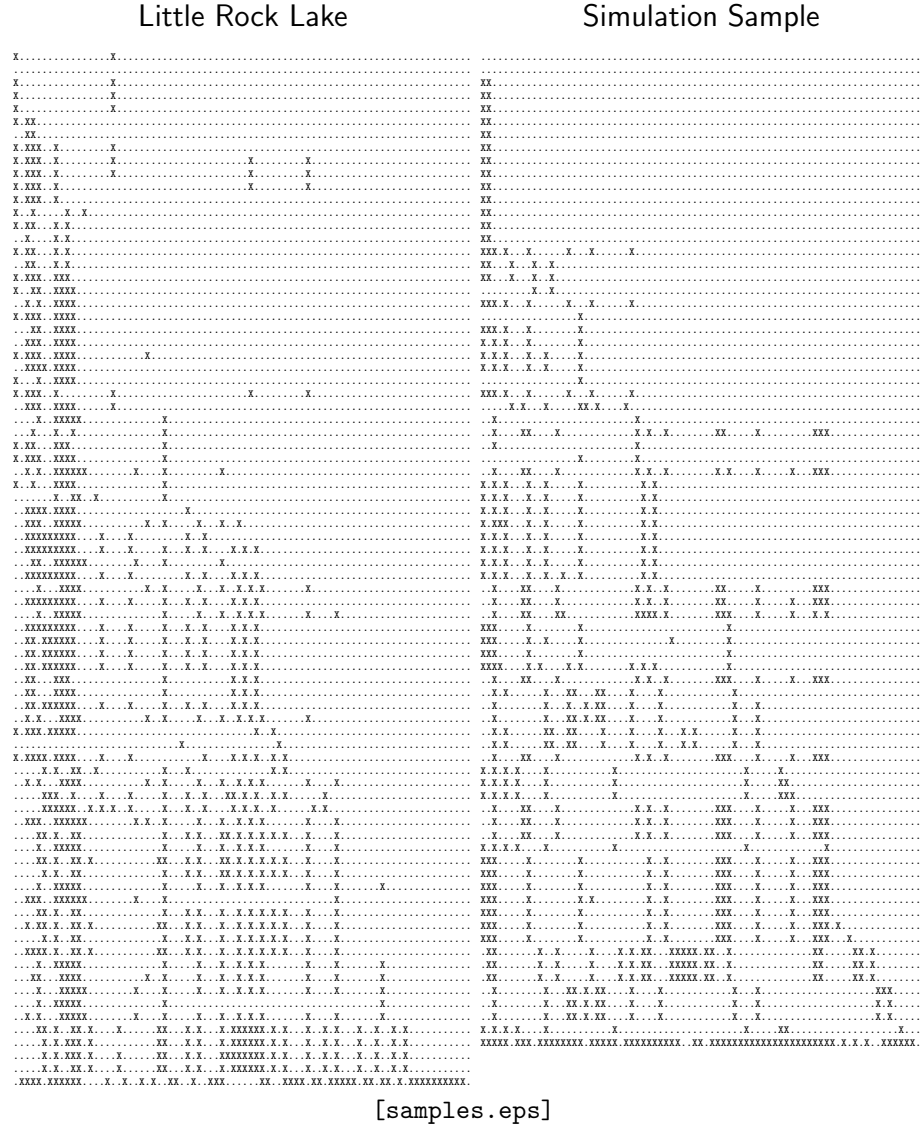
[samples.eps]

Figure 2: Examples for connection matrices of food webs

The connection matrix of the food webs of Little Rock Lake (Martinez, 1991) and a simulation sample, both after species lumping (see Appendix B for lumping procedure). Model parameters are as in Table 5, Little Rock. Each 'X' indicates that the species corresponding to the column eats the species corresponding to the row. The ordering of the species is such as to minimize the number of upper diagonal links and otherwise random.
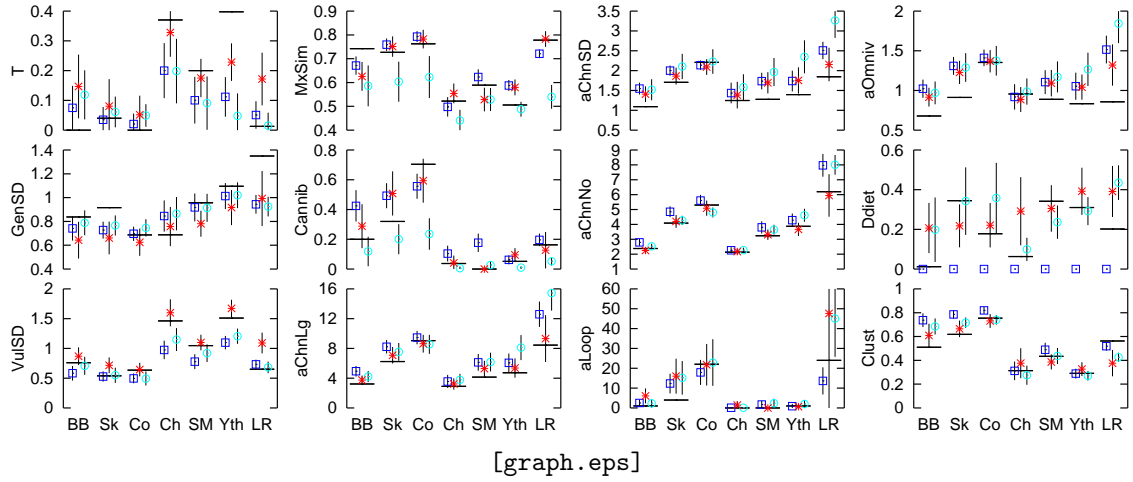
Figure 3: Comparison of model and empirical data

The graphs display the values of twelve food-web statistics (Appendix B) of seven food webs (key in Table 5) obtained from the speciation model (red stars), in comparison with the empirical data (horizontal lines), the niche model (Williams and Martinez, 2000) (blue squares), and the nested hierarchy model (Cattin et al., 2004) (blue circles). Vertical lines indicate model standard deviations.